

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 10.1508 Seconds
(without alignments)
1581.285 Million cell updates/sec

Title: US-09-807-933B-9

Perfect score: 2106

Sequence: 1 MKFTVAITSAVALSSA.....TFKEVTCPAELTTRSGCERK 387

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|--------------|--------------------|
| 1 | 714 | 33.9 | 376 | 1 GUNK_FUSOX | P45699 fusarium ox |
| 2 | 710 | 33.7 | 213 | 1 GUN5_HUMIN | P43316 humicola in |
| 3 | 592 | 28.1 | 511 | 1 GUNB_PSEFL | P18126 pseudomonas |
| 4 | 525.5 | 25.0 | 393 | 1 GUN1_USTMA | P54424 ustilago ma |
| 5 | 226.5 | 10.8 | 471 | 1 GUX2_TRIRE | P07987 trichoderma |
| 6 | 221 | 10.5 | 210 | 1 P8SP_PORPU | P50272 porphyra pu |
| 7 | 190 | 9.0 | 418 | 1 GUN2_TRIRE | P07982 trichoderma |
| 8 | 177 | 8.4 | 475 | 1 SIM1_YEAST | P40472 saccharomyc |
| 9 | 174.5 | 8.3 | 378 | 1 SLG1_YEAST | P54867 saccharomyc |
| 10 | 172 | 8.2 | 438 | 1 GUX3_AGABI | P49075 agaricus bi |
| 11 | 166.5 | 7.9 | 462 | 1 GUNB_FUSOX | P46236 fusarium ox |
| 12 | 166.5 | 7.9 | 481 | 1 LORI_MOUSE | P18165 mus musculu |
| 13 | 166.5 | 7.9 | 503 | 1 YN23_YEAST | P53832 saccharomyc |
| 14 | 166 | 7.9 | 937 | 1 HYR1_CANAL | P46591 candida alb |
| 15 | 163 | 7.7 | 600 | 1 SP96_DICDI | P14328 dictyosteli |
| 16 | 161.5 | 7.7 | 662 | 1 MUC1_XENLA | Q05049 xenopus lae |
| 17 | 159.5 | 7.6 | 2715 | 1 G156_PARPR | P13837 paramesium |
| 18 | 158.5 | 7.5 | 537 | 1 SP70_DICDI | P15269 dictyosteli |
| 19 | 158 | 7.5 | 786 | 1 STUB_DROME | Q05319 drosophila |
| 20 | 158 | 7.5 | 1367 | 1 AMTH_YEAST | P08640 saccharomyc |
| 21 | 156.5 | 7.4 | 2704 | 1 G168_PARPR | P17053 paramesium |
| 22 | 153.5 | 7.3 | 385 | 1 GUNF_FUSOX | P46239 fusarium ox |
| 23 | 153 | 7.3 | 242 | 1 GUN5_TRIRE | P43317 trichoderma |
| 24 | 151.5 | 7.2 | 962 | 1 GUNA_PSEFL | P10476 pseudomonas |
| 25 | 151 | 7.2 | 1246 | 1 YWV2_CABEL | P34504 caenorhabdi |
| 26 | 148.5 | 7.1 | 389 | 1 SER1_BOMMO | P07856 bombyx mori |
| 27 | 146 | 6.9 | 542 | 1 SCWB_YEAST | P53189 saccharomyc |
| 28 | 146 | 6.9 | 605 | 1 YHC8_YEAST | P38739 saccharomyc |
| 29 | 145.5 | 6.9 | 1150 | 1 APNU_PIG | P12021 sus scrofa |
| 30 | 145 | 6.9 | 1218 | 1 JAG1_MOUSE | Q98400 mus musculu |
| 31 | 144.5 | 6.9 | 1408 | 1 SERR_DROME | P18168 drosophila |
| 32 | 144 | 6.8 | 513 | 1 GUX1_TRIRE | P00725 trichoderma |
| 33 | 144 | 6.8 | 686 | 1 DLL4_MOUSE | Q9ji71 mus musculu |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 34 | 144 | 6.8 | 757 | 1 VGLX_HSVBB | P28968 equine herp |
| 35 | 142.5 | 6.8 | 555 | 1 DP87_DICDI | Q04503 dictyosteli |
| 36 | 142.5 | 6.8 | 573 | 1 C114_MOUSE | P19467 mus musculu |
| 37 | 142 | 6.7 | 513 | 1 GUX1_TRIVI | P19355 trichoderma |
| 38 | 142 | 6.7 | 1218 | 1 JAG1_HUMAN | P78504 homo sapien |
| 39 | 141.5 | 6.7 | 1200 | 1 ICEN_PSESY | P06620 pseudomonas |
| 40 | 141.5 | 6.7 | 1224 | 1 PER_DROME | P07663 drosophila |
| 41 | 141 | 6.7 | 1561 | 1 PERB_HUMAN | Q15648 h peroxisom |
| 42 | 141 | 6.7 | 2471 | 1 NTC2_HUMAN | Q04721 homo sapien |
| 43 | 140.5 | 6.7 | 459 | 1 GUN1_TRIRE | P07981 trichoderma |
| 44 | 140.5 | 6.7 | 540 | 1 GUX1_ASPAC | O59843 aspergillus |
| 45 | 139.5 | 6.6 | 463 | 1 GUN1_TRILO | Q12714 trichoderma |

ALIGNMENTS

RESULT 1

| ID | GUNK_FUSOX | STANDAED | PRT | 376 AA |
|----|---|----------|-----|--------|
| AC | P45699; | | | |
| DT | 01-NOV-1995 (Rel. 32, Created) | | | |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update) | | | |
| DT | 01-NOV-1995 (Rel. 32, Last annotation update) | | | |
| DE | Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase). | | | |
| OS | Fusarium oxysporum. | | | |
| OC | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; | | | |
| OC | Hypocreales; mitosporic Hypocreales; Fusarium. | | | |
| OX | NCBI_TaxID=5507; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=95047531; PubMed=7959045; | | | |
| RA | Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C., Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.; | | | |
| RT | "The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from Fusarium oxysporum."; | | | |
| RL | Gene 150:163-167(1994). | | | |
| CC | -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose. | | | |
| CC | -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD). | | | |
| CC | -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL HYDROLASES). | | | |

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

| | |
|---|--------------------------------|
| EMBL; L29381; AAA65589.1; -- | |
| HSSP; P43316; 2ENG. | |
| InterPro; IPR000254; CBD_fungal. | |
| InterPro; IPR000334; GH_45. | |
| Pfam; PF00734; CBM_1; 1. | |
| Pfam; PF02015; Glyco_hydro_45; 1. | |
| SMART; SM00236; fCBD; 1. | |
| PROSITE; PS00562; CBD_FUNGAL; 1. | |
| PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1. | |
| KW Cellulose degradation; Hydrolase; Glycosidase; Signal. | |
| FT SIGNAL 1 18 | POTENTIAL. |
| FT CHAIN 19 376 | PUTATIVE ENDOGLUCANASE TYPE K. |
| FT DOMAIN 19 308 | CATALYTIC. |
| FT DOMAIN 309 338 | LINKER. |
| FT DOMAIN 339 376 | CELLULOSE-BINDING. |
| FT ACT SITE 29 29 | NUCLEOPHILE (BY SIMILARITY). |
| FT ACT SITE 140 140 | PROTON DONOR (BY SIMILARITY). |
| SQ SEQUENCE 376 AA; 39235 MW; B430A5F962B9F882 CRC64; | |

Query Match 33.9%; Score 714; DB 1; Length 376;
Best Local Similarity 59.1%; Pred. No. 2.8e-38;

[illegible]

| RESULT 2 | GUNS | HUMIN | STANDARD; | PRT; | 213 AA. |
|----------|--|--------------------|-----------|------|---------|
| ID | GUNS | HUMIN | STANDARD; | PRT; | 213 AA. |
| AC | P43316; | | | | |
| DT | 01-NOV-1995 (Rel. 32, Created) | | | | |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update) | | | | |
| DT | 15-DEC-1998 (Rel. 37, Last annotation update) | | | | |
| DE | Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V) | | | | |
| DE | (Cellulase V) (EG V). | | | | |
| OS | Humicola insolens. | | | | |
| OC | Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola. | | | | |
| OC | NCBI_TaxID=34413; | | | | |
| RN | [1] | SEQUENCE FROM N.A. | | | |
| RP | Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F., | | | | |
| RP | Hjort C.M., Hastrup S.; | | | | |
| RA | "A cellulase preparation comprising an endoglucanase enzyme."; | | | | |
| RA | Patent number WO9117243, 14-NOV-1991. | | | | |
| RL | [2] | | | | |
| RN | X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS). | | | | |
| RP | MEDLINE=93390621; PubMed=8377830; | | | | |
| RX | Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z., | | | | |
| RA | Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulein M.; | | | | |
| RA | "Structure and function of endoglucanase V."; | | | | |
| RL | Nature 365:362-364(1993). | | | | |
| RL | [3] | | | | |
| RN | X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS). | | | | |
| RP | MEDLINE=96101453; PubMed=8519779; | | | | |
| RX | Davies G.J., Tolley S.P., Henriessat B., Hjort C., Schulein M.; | | | | |
| RA | "Structures of oligosaccharide-bound forms of the endoglucanase V | | | | |
| RT | from Humicola insolens at 1.9-A resolution."; | | | | |
| RT | Biochemistry 34:16210-16220(1995). | | | | |
| RL | [4] | | | | |
| RN | X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS). | | | | |
| RP | Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z., | | | | |
| RA | Wilson K.S., Rasmussen G., Schulein M.; | | | | |
| RA | "Structure determination and refinement of the Humicola insolens | | | | |
| RT | endoglucanase V at 1.5-A resolution."; | | | | |
| RT | Acta Crystallogr. D 52:7-17(1996). | | | | |
| RL | -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic | | | | |
| CC | linkages in cellulose. | | | | |
| CC | -!- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOS | | | | |
| CC | HYDROLASES). | | | | |
| CC | PDB; 2ENG; 08-DEC-96. | | | | |
| DR | PDB; 3ENG; 16-JUN-97. | | | | |
| DR | PDB; 4ENG; 16-JUN-97. | | | | |
| DR | InterPro; IPR000334; GH 45. | | | | |
| DR | Pfam; PF02015; Glyco hydro 45; 1. | | | | |
| DR | PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1. | | | | |
| KW | Cellulose degradation; Hydrolase; Glycosidase; 3D-structure. | | | | |
| FT | ACT_SITE 10 10 NUCLEOPHILE. | | | | |
| FT | ACT_SITE 121 121 PROTON DONOR. | | | | |
| FT | SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64; | | | | |

```

Query Match          33.7%; Score 710; DB 1; Length 213;
Best Local Similarity 62.1%; Pred.No. 3e-38;
Matches 126; Conservative 25; Mismatches 46; Indels 6; Gaps 3;

QY      186   GSTRYWDCCKASCSPFKASVTPGVDTCSANGISLLDANAQSCN-GGNGFMCNNQPW 244
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db       3    GRSTRYWDCCKPCGWAKKAPVNPQVFSCNANFORITDFDAKSGCEPGGVAYSCADOTPW 62
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY      245   AVNDELAYGFAAASIAGSNEAGCCGCYELTFTSGAASGKKMWVVQTNTGGDLGSNHFDL 304
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db       63   AVNDPFLGPAANTSIAGSNEAGWCACCYELTFTSGPVAGKKMWWQVSTTGGDLGSNHFDL 122
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY      305   QMPGGVGIFENGCAAQWGA-PNDGWGARYGGVSVDPSALPSALQAGCKWRFWFKNSD 363
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db      123   NIPGGVGIFEDGCTPOFGGLP---GQRYGGISRNECDRFPDALKPGCYWRPWFKNAD 178
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY      364   NPWTPTKEVTCPAELTTTRSCGR 386
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db      179   NPSFSFRQVCQPAELVARTGRR 201
         |||:|||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

RESULT 3
GUNB_PSEFL STANDARD; PRT; 511 AA.
ID   GUNB_PSEFL AC P8126;
DT   01-NOV-1990 (Rel. 16, Created)
DD   01-NOV-1990 (Rel. 16, Last sequence update)
DR   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DS   (Cellulase) (EGB).
OS   CELB.
GN   Pseudomonas fluorescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=294;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
RC   MEDLINE=90355836; PubMed=2117693;
RX   Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
RA   "The N-terminal region of an endoglucanase from Pseudomonas
RT   fluorescens subspecies cellulosa constitutes a cellulose-binding
RL   domain that is distinct from the catalytic centre.";
RL   Mol. Microbiol. 4:759-767(1990).
CC   -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOTHEROLYSIS OF 1,4-BETA-
CC   GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC   GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED
CC   SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.
CC   -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC   linkages in cellulose.
CC   -1- SUBCELLULAR LOCATION: Periplasmic.
CC   -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC   (CBD).
CC   -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC   HYDROLASES).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a licensee agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X52615; CAA36844.1; -.
DR   PIR; S10527; S10527.
DR   HSSP; P43316; 2ENG.
DR   InterPro; IPR001919; Bac cellose-bind.
DR   InterPro; IPR002883; CBD_5.
DR   InterPro; IPR000334; GH_45.
DR   Pfam; PF00553; CBM 2; 1.
DR   Pfam; PF02013; CBM 10; 1.
DR   Pfam; PF02015; Glyco hydro 45; 1.

```


[2] SEQUENCE FROM N.A.
RC STRAIN-QM9414 / Rut C-30;
RA Chen C.M., Gritzali M., Stafford D.W.;
RT "Nucleotide sequence and deduced primary structure of
RL cellobiohydrolase II from Trichoderma reesei";
RN Biotechnology 5:274-278(1987).
[3]
SEQUENCE OF 25-44.
RP Faegerstaam L.G., Pettersson L.G.;
RA "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei QM
RT 9414";
RL FEBS Lett. 119:97-100(1980).
RN [4]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RP MEDLINE=90333255; PubMed=2377893;
RX Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
RA "Three-dimensional structure of cellobiohydrolase II from Trichoderma
RT reesei.";
RL Science 249:380-386(1990).
RN [5]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RP MEDLINE=97029636; PubMed=8875646;
RX Koivula A., Reinikainen T., Ruohonen L., Valkeajärvi A.,
RA Claessens M., Telman O., Kleywegt G.J., Szardenings M., Rouvinen J.,
RA Jones T.A., Teeri T.T.;
RT "The active site of Trichoderma reesei cellobiohydrolase II: the role
of tyrosine 169";
RL Protein Eng. 9:691-699(1996).
CC -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOLGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOHYDRALASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLIZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: T. REESEI PRODUCES TWO DIFFERENT
CC EXOCELLULOHYDRALASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYZE
CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOLGLUCANASES.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licens@isb-sib.ch).
CC -----
DR EMBL; M16190; AAA34210.1; --
DR ENBL; M55080; AAA72922.1; --
DR PIR; A26472; A26472.
DR PDB; A26160; A26160.
DR PDB; 1CB2; 19-MAR-99.
DR InterPro; IPR000254; CBD fungal.
DR Pfam; PF00734; CBM_1; 1.
DR PRINTS; PF01341; Glyco hydro 6; 1.
DR PRODOM; PD00733; GLHYDLASE6.
DR ProDom; PD001821; CBD fungal; 1.
DR SMART; SM00236; fcbd; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative polysaccharide binding protein precursor.
OS Porphyra purpurea.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Liu Q., der Meer J.P., Reith M.E.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 4 FUNGAL-TYPE CELLULOSE-BINDING DOMAINS
CC (CBD).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U08843; AAA61792.1; -;
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR Pfam; PF00734; CBM_1; 4.
DR SMART; SM00236; fCBD; 4.
DR PROSITE; PS00562; CBD_FUNGAL; 4.
KW Repeat; Signal.
FT SIGNAL 1 22
FT CHAIN 23 210 PUTATIVE POLYSACCHARIDE BINDING PROTEIN.
FT DOMAIN 23 62 POLYSACCHARIDE BINDING DOMAIN I.
FT DOMAIN 63 105 POLYSACCHARIDE BINDING DOMAIN II.
FT DOMAIN 125 165 POLYSACCHARIDE BINDING DOMAIN III.
FT DOMAIN 166 210 POLYSACCHARIDE BINDING DOMAIN IV.
SQ SEQUENCE 210 AA; 21967 MW; 83782D98B6863CC8 CRC64;

Query Match 10.5%; Score 221; DB 1; Length 210;
Best Local Similarity 29.1%; Pred. No. 1.4e-07;
Matches 69; Conservative 22; Mismatches 74; Indels 72; Gaps 11;

Qy 11 AVALALSSAAAS-CSSVYQCGIGWTGPTCCDAGSTCKAQDKNKYSQC--IPKPKG 67
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 65
8 AAALTLSAAAASACGVLYEQCGIGDGTCCSEGLMC--MKMGPIYSQCRAMPGMWG 65
Qy 68 SSSSSSSSVSYQCGIGWGPCTCESGSTCVAQEGNKYSQCLPGSHNNAGNASSTKK 127
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 108
66 QVKE-----YQCGGMNYSKWTWCSPGFKCV--ELNEFFSQ-----DLANKSPV 108
Qy 128 TSTKTSTTTAKATATVTK 187
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 141
109 ATPKVPSTPPGPAQVCKEY-----AACGGEMFNGAK----- 141
Qy 188 TTRYWDCC-----ASCSPFGKASVTPVDTCASNGISLLDNAQSGCNGGFM 237
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 182
142 -----CKGFLGVCVETSGKWSQCRAPP-----KMGVCRYAQCGNGYM 182

RESULT 7
GUN2 TRIE
ID_GUN2 TRIE STANDARD; PRT; 418 AA.
AC P07982;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
DE Endoglucanase EG-II precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN EGL2 OR EGLI.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VTT-D-80133;
RX MEDLINE=98255850; PubMed=3384334;
RA Saloheimo M., Lehtoavaara P., Penttilae M., Teeri T.T., Staahlberg J.,
RA Johansson G., Pettersson G., Clayssens M., Tomme P., Knowles J.K.C.;
RT "EGLII, a new endoglucanase from Trichoderma reesei: the
RL characterization of both gene and enzyme.";
RL Gene 63:11-21(1988).
RN [2]
RP ACTIVE SITE GLU-350.
RX MEDLINE=93131031; PubMed=8093602;
RA Macarion R., van Beeumen J., Henrissat B., de la Mata I.,
RA Claeysens M.;
RT "Identification of an essential glutamate residue in the active site
RL of endoglucanase III from Trichoderma reesei.";
RL FEBS Lett. 316:137-140(1993).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIODOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- CAUTION: WAS ORIGINALLY CALLED ENDOGLUCANASE EG-III.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; M19373; AAA34213.1; -;
DR PIR; S28372; S28372.
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00734; CBM_1; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 21
FT CHAIN 22 418 ENDOGLUCANASE EG-II.
FT DOMAIN 22 57 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 58 91 LINKER.
FT DOMAIN 92 418 CATALYTIC.
FT MOD RES 22 22 PYRROLIDONE CARBOXYLIC ACID
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 29 46 BY SIMILARITY.
FT DISULFID 40 56 BY SIMILARITY.
FT ACT SITE 239 239 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 350 350 NUCLEOPHILE.
SQ SEQUENCE 418 AA; 44227 MW; 26A492D55237A49B CRC64;

Query Match 9.0%; Score 190; DB 1; Length 418;
Best Local Similarity 24.3%; Pred. No. 2.2e-05;
Matches 83; Conservative 41; Mismatches 119; Indels 98; Gaps 15;

Qy 76 SVYQCGIGWGPCTCESGSTCVAQEGNKYSQCLPGSHNNAGNASSTKTKTSTT 135
Dy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 60
24 TWVQCGIGWGPCTNCPAGSACSTL--NPYYAQCFGA----- 60
Qy 136 TAKATATVTK 195

```

Db 61  ---TTTITTPRSGTTRATSTSSPTSSGVRPAGVNIAGFDGCTTD--GTC 113
Qy 196 KASCSWPQKASVGTVPDTCAN-----GSLLDANAQSCNGGFMCMNNPFWAVNDELA 251
Db 114 VTSKYVPLKNTG-----SNNYPDGIQGMQHFVNE--DGMTIFRLPVGMQVLYVNNLIG 165
Qy 252 YGFAAASAGSNEAGWCCGCVELFTSAAAGKMMVQVTN-----TGGDLGSHNFDLQMP 307
Db 166 GNLSTSTISKYDQL--VQGCISL-----GAYCIVDIHNYRWNGGIIGQ----- 207
Qy 308 GGG-----VGIFNGCAQWGA-----PND-----GWGARYGGV-----SSVS 339
Db 208 -GGPNAQFTSLMSQASQSYASQSRVFGWIFWNEPHDVNINWTAATVQVVTAINRAGATS 266
Qy 340 DCASLPALQAGCKRWFRNFKNSDNTMTTFKEVTCPAELTT 380
Db 267 QFISLP-----GNDWQAGAFISDGAASAAALSQVTPNPDGSTT 302

```

RESULT 8

```

SIM1 YEAST
ID SIM1 YEAST STANDARD; PRT; 475 AA.
AC P40472;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SIM1 protein precursor.
GN SIM1 OR PBP3 OR YIL123W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

```

SEQUENCE FROM N.A.

```

RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horenell T.S., Hunt S., Jagels K., Jones M.,
RA Louis B., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]

```

IDENTIFICATION.

```

RX MEDLINE=9615611; PubMed=8574583;
RA Dahmann C., Diffley J.F.X., Nasmyth K.A.;
RT "S-phase-promoting cyclin-dependent kinases prevent re-replication by
RT inhibiting the transition of replication origins to a pre-replicative
RT state.";
RL Curr. Biol. 5:1257-1269 (1995).
CC -!- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE CLB5 KINASE
CC ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE SUN FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46833; CAA86869.1; -
DR SGD; S0001385; SIM1.
KW SIGNAL.
FT SIGNAL. 1 19 POTENTIAL.
FT CHAIN 20 475 SIM1 PROTEIN.
FT DOMAIN 58 112 ALA-RICH.
FT DOMAIN 80 203 SER-RICH.
FT CAROXYD 422 422 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 475 AA; 48070 MW; B7B0F0BC8478612F CRC64;

```

```

Query Match 8.4%; Score 177; DB 1; Length 475;
Best Local Similarity 24.6%; Pred. No. 0.00016;
Matches 98; Conservative 51; Mismatches 139; Indels 110; Gaps 18;

Qy 4 TVAITTSIALALSSSAAASCSVYQCGGIGWTGPTCCDAGSTCKAQNKY----YS 59
Db 67 TSAIITTSV-LAPTSSAAAGIAASI-----AVSSAALAKNEKISDRAAS 109
Qy 60 QCIPKPKGSSSSSSCSVYQCGGIGWGTCCBSGTCVAQEGNKYISQCLPQSHSNA 119
Db 110 ATASTSQGASSSSSSSSATS-----TLESSTV-----SSSSSEA 143
Qy 120 GNASSTKTKSTSTTTAKAT--ATVTKTVTKTTTSTTAAASTSTSSAGYKV 177
Db 144 APTSTVVSTTSATOSSASSATKSTSTSTSTSTSTSTSTSTSTSTSTSTSTSSSSSSSSSSSSGGS 203
Qy 178 ISGGKSG-SGSTTRYMD---CCK-----ASCWPKGKASVTPVDTCAASNGISLIDAN 225
Db 204 IYGLADFSGPSEKFDQGTIPCDKFPSSQGVISIDWIEGGWGSVENTDTSTGSCKEGS 263
Qy 226 AQS-GCNGG-----NGFMNNOQPWAVNDELA-----GFAAASIAQ-- 261
Db 264 YCSYSCQPGMSKTOWPSDQPSDGRSVGGLLCKNGLYRSNTDADYLCWGVAAAYVWSKL 323
Qy 262 SNEAGWCCGYE-----LTFTSQAASGKKMVVQ-----VTNTGDLGSHNFDLQMPGG 310
Db 324 SKGVAICRTDYPGTENMVIPTVEGSSSLPLTVVDQDTYFTWEGKTKSAQYI---VNNAG 380
Qy 311 VGIFNGCAQWGPNDGNG-----ARYGGVSSVS 339
Db 381 VSVEDGCI--WGTSGSGIGNWAPLNFAGAGTGGVTYLS 416

RESULT 9
SIM1 YEAST
ID SIM1 YEAST STANDARD; PRT; 378 AA.
AC P54857;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SLG1 protein precursor.
GN SLG1 OR YOR008C OR UNF378.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RA de Bettignies G., Bergez-Aullo P., Barthe C., Louvet O.,
RA Peypouquet M.F., Morel C., Doignon F., Crouzet M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=9705159; PubMed=8896276;
RA Starky F., Holmberg A., Pettersson B., Uhlen M.;
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
RT from Saccharomyces cerevisiae reveals 15 open reading frames, five of
RT which correspond to previously identified genes.";
RL Yeast 12:1091-1095 (1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39481; AAA85862.1; -
DR EMBL; U43491; AAC49488.1; -
DR EMBL; Z74916; CAA99196.1; -
DR SGD; S0005534; SLG1.
DR InterPro; IPR002889; WSC.

```


Eukaryota; fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; mitosporic Hypocreales; Fusarium.

RP SEQUENCE FROM N.A.
RA MEDLINE=95047531; PubMed=7959045;
RA Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
RA "The use of conserved cellulase family-specific sequences to clone
RT cellulase homologue cDNAs from Fusarium oxysporum.";
RL Gene 150:163-167(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L29377; AAA65595.1; -
CC HSP; F07987; ICB2.
CC InterPro; IPR000254; CBD_fungal.
CC InterPro; IPR001524; GH_6.
CC Pfam; PF00734; CBM_1; 1.
CC Pfam; PF01341; Glyco_hydro_6; 1.
CC PRINTS; PR00733; GLYDRLASE6.
CC ProDom; PD001821; CBD_fungal; 1.
CC ProDom; PD003733; GH_6; 1.
CC SMART; SM00236; fCBD; 1.
CC PROSITE; PS00562; CBD_FUNGAL; 1.
CC PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
CC PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
KW SIGNAL 1 16
FT CHAIN 17 462 PUTATIVE ENDOGLUCANASE TYPE B.
FT DOMAIN 17 65 CELLULOSE-BINDING.
FT DOMAIN 66 99 LINKER.
FT DOMAIN 100 462 CATALYTIC.
FT ACT_SITE 190 190 BY SIMILARITY.
FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 416 416 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 33 50 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 191 250 BY SIMILARITY.
FT DISULFID 383 430 BY SIMILARITY.
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 462 AA; 49207 MW; E25B2F5B828B637F CRC64;

Query Match 7.9%; Score 166.5; DB 1; Length 462;
Best Local Similarity 44.3%; Pred. No. 0.00071;
Matches 43; Conservative 9; Mismatches 26; Indels 19; Gaps 4;

QY 73 SCSS-VYSCGGTGWGGTCCSGTCAQEGNKYSQCLPGS--HSNAGNASSTTKTS 129
Db 24 SCNGVWACGGQWNGTTCCTSGNKC--KLNDFYSQCQPGSAEPSSTAAGPSST---- 77

QY 130 TKTSTTTAKATAVTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 166
Db 78 -----TATKTTATGSSSTTAGSVTSAPPAAS 104

RESULT 12
LORI MOUSE STANDARD; PRT; 481 AA.
AC P18165;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT DE 15-JUN-2002 (Rel. 41, Last annotation update)
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90275605; PubMed=2190691;
RA Mehrel T., Hohl D., Rothnagel J.A., Longley M.A., Bundman D.,
RA Cheng C., Lichi U., Bisher M.E., Steven A.C., Steinart P.M.,
RA Yuspa S.H., Roop D.R.;
RT "Identification of a major keratinocyte cell envelope protein,
RT loricrin.";
RL Cell 61:1103-1112(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=95256248; PubMed=7738016;
RA Diserio D., Jones A., Bundman D., Rothnagel J.A.,
RA Roop D.R.;
RT "The proximal promoter of the mouse loricrin gene contains a
RT functional AP-1 element and directs keratinocyte-specific but not
RT differentiation-specific expression.";
RL J. Biol. Chem. 270:10792-10799(1995).
CC -1- FUNCTION: MAJOR KERATINOCYTE CELL ENVELOPE PROTEIN.
CC -1- SUBUNIT: MONOMERS ARE CROSSLINKED BY DISULFIDE AND N-(GAMMA-
CC GLUTAMYL) LYSINE ISODIPEPTIDE BONDS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M34398; AAA39444.1; -
CC EMBL; U09189; AAA82152.1; -
CC PIR; A35628; A35628.
CC HSP; P02876; 9WGA.
CC MGD; MG1:96816; Lor.
KW Keratinization.
SQ SEQUENCE 481 AA; 37830 MW; 97349A786FF239FE CRC64;

Query Match 7.9%; Score 166.5; DB 1; Length 481;
Best Local Similarity 24.0%; Pred. No. 0.00074;
Matches 83; Conservative 19; Mismatches 121; Indels 123; Gaps 16;

QY 19 SAEAAACSSVY-----QCQGGIG---WTGPTCCDAGSTCKAKDNKYSSCIPKPKGSSS 70
Db 222 SSGSSCGGGYSGGGSSCGGGYSGGGTSCGGSSGGGGSSGQYQCQYGGGSSG 281
QY 71 SSSCSSVY-----SQCQG--IGWSGPTC---CBSGTCVAQEGNKY---SQCLPESH 115
Db 282 GSSCGGYSGGGSSCGGYSGGGSSCGGGSSCGGGSSCGGGSSCGGGSSCGGGSS 341
QY 116 SNNAGNASSTTKTSKTKSTTTAKATAVTTKTKTKTKTKTKTKTKTKTKTKTK 175
Db 342 GGGGGYSSQ-----TSQTSAPQ 361
QY 176 KVISGGKSGSTTRYWDCKKASCWPKASVTGPVDTCASNGISLLDANAQSCNGNG 235
Db 362 QSYGGSSGGG-----GSCG--GGSSGGGGGGCYSSG-----GGSSGGCGGG-- 403
QY 236 FMCNNPQWVNDLAYGFAAASIAGSNEAGWCCGCVYELFTSAAAGSKMVOVTNFG 295
Db 404 -----YSGGGGCGGGSSGGSGGC-----GGSSG-----GSGG 433
QY 296 DLGNSHFDLPMPGGVGIFNCAQWGPNDGWARGYGVSSVSDC 341
Db 434 GCGGGY-----SGG-----GGSSCGGGSSGGSS--GGKGVPVC 467

FT CARBOHYD 619 619 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 641 641 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 913 913 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 937 AA; 93699 MW; 17802E12E5BA926 CRC64;

Query Match 7.9%; Score 166; Length 937;
Best Local Similarity 23.6%; Pred. No. 0.0014;
Matches 111; Conservative 49; Mismatches 136; Indels 174; Gaps 20;

QY 6 AITSAVALALSSAASCSVYQCGIGWT--GPTCCDAGSTCAQKDNKYYSQCIP 63
Db 372 SLTSSSLSKSESSDVVSTTNISSSTAIETWNSESTDAGSSISQSSESS--STAIT 429
QY 64 KPGKSSSSSSCSVYQCGIGWGPTCCESGTCVAQEGNKYYSQCLPGSHSNAGNA- 122
Db 430 SSETSSSESMSASSTTASN-----TSIETDSGIVSQ-----SESSNAL 469
QY 123 SSTKT-----STKTSTTTA---KATATVTT----- 145
Db 470 SSTEQSISSPGQSTIYNSTVTSITSCDENKCTEDVVTFTVPCSTDCVPTTGDI 529
QY 146 -----KVTYKTTK-----TTKTSTTAASSTSSAGYKVISGK- 182
Db 530 STSYQTQVTTITNCDEVSCQDVVYVTVNPHVTVDATTTTSTGDNSTGNEGS 589
QY 183 -----SGSSTTRYWDCCAKSCWPKASVGPVDTCASNGISLLDANAAQ 227
Db 590 NHGPNGSTEGSGNGSGAGSN-----EGSQSGPNNGSGG-----SEGSNNGSGD 636
QY 228 SGCGNGGFMCMN-----NQPWA-VNDELAYGFMAASIAAG 261
Db 637 SGSNNGSGSGNNGSGSGSTEGSGNEGSQSGSQPGPNEGSGSGSGNEGSNHG 696
QY 262 SNEA-----GWCCCYELFTSCAAGKVMVQV-TNTGDLGSHNFDLQMPGG 309
Db 697 SNEGSGSGSGNNGSGSGS-----QSGSGSGSGSGSGSGSGSGNNGSGN-----PGA 745
QY 310 GVCIFNGC-----AAQWAPNDGWARYG-GVSSVSDCASLPFA 347
Db 746 GNGSGSGSGSGSGSGSGAGSGSGGSGPNNGSGSHNDGSGSGSGNPGCA 795

RESULT 15
SP96 DICDI STANDARD; PRT; 600 AA.
AC P14378;
DT 01-JAN-1990 (Rel. 13, Created).
DT 01-JAN-1990 (Rel. 13, Last sequence update).
DT 01-NOV-1997 (Rel. 35, Last annotation update).
DE Spore coat protein SP96.
GN COTA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=90067962; PubMed=2587278;
RA Feenagh K., Loomis W.F.;
RT "Sequence of the Dictyostelium discoideum spore coat gene SP96.";
RL Nucleic Acids Res. 17:9489-9489(1989).
CC -1- SUBCELLULAR LOCATION: OUTER LAYER OF THE COAT MATRIX AND
CC INTERSPORE MATRIX.
CC -1- PTM: PHOSPHORYLATED AND FUCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED,
CC MAY CONTAIN GLCNAC-ALPHA-1-P-SER RESIDUES.
CC -1- SIMILARITY: CONTAINS 4 PRESPORE MOTIFS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16491; CAA34508.1; -;
DR PIR; S07638; S07638;
DR GlycosuiteDB; P14328; -;
DR DictyDB; DDO3007; COTA.
DR InterPro; IPR003645; FOLN.
DR SMART; SM00274; FOLN; 6.
KW Glycoprotein; Phosphorylation; Repeat; Sporulation.
FT REPEAT 185 197 PRESPORE MOTIF 1.
FT REPEAT 221 233 PRESPORE MOTIF 2.
FT REPEAT 298 310 PRESPORE MOTIF 3.
FT REPEAT 395 407 PRESPORE MOTIF 4.
SQ SEQUENCE 600 AA; 59589 MW; 616AE6D02B5F1071 CRC64;

Query Match 7.7%; Score 163; DB 1; Length 600;
Best Local Similarity 26.3%; Pred. No. 0.0015;
Matches 49; Conservative 35; Mismatches 92; Indels 10; Gaps 2;

QY 4 TVAITSAVALALSSAASCSVYQCGIGWTGPTCCDAGSTCAQKDNKYYSQCIP 63
Db 422 TTGSTSDSSALGSTSESSASGSAVSSA-----SGSSAASSPSSSAASSPSSSAASS 476
QY 64 KPGKSSSSSSCSVYQCGIGWGPTCCESGTCVAQEGNKYYSQCLPGSHSNAGNAS 123
Db 477 SPSSAASSPSSSAASS-----SPSSAASSPSSSAASSPSSSAASSPSSSAASS 531
QY 124 STKTSTKTSTTTAKATATVTTKTSTTTTAAATSTSSAGYKVISGKXS 183
Db 532 SSAATTAATTATTAATTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 591
QY 184 GSGSTT 189
Db 592 TTATTT 597

Search completed: June 18, 2003, 15:32:07
Job time : 11.1508 secs